PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : G10L 19/00

A2

(11) International Publication Number: WO 00/45378

(43) International Publication Date: 3 August 2000 (03.08.00)

(21) International Application Number: PCT/SE00/00158

(22) International Filing Date: 26 January 2000 (26.01.00)

(30) Priority Data:

9900256-0	27 January 1999 (27.01.99)	SE
9903552-9	1 October 1999 (01.10.99)	SE

(71)(72) Applicant and Inventor: LILJERYD, Lars, Gustaf [SE/SE]; Vintervägen 19, S-171 34 Solna (SE).

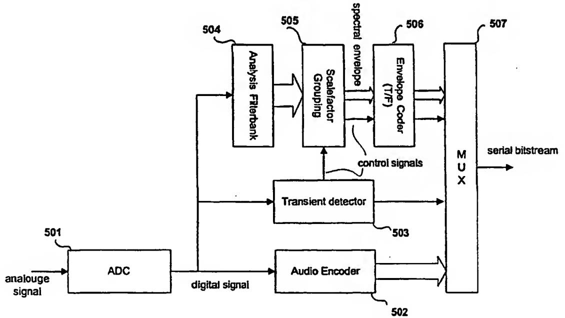
(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: KJÖRLING, Kristofer [SE/SE]; Lostigen 10, S-170 75 Solna (SE). EKSTRAND, Per [SE/SE]; Renstiernas Gata 29, S-116 31 Stockholm (SE). HENN, Fredrik [SE/SE]; Ritarvägen 14, S-168 31 Bromma (SE).

(74) Agents: ÖRTENBLAD, Bertil et al.; Noréns Patentbyrå AB, Box 10198, S-100 55 Stockholm (SE).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: EFFICIENT SPECTRAL ENVELOPE CODING USING VARIABLE TIME/FREQUENCY RESOLUTION AND TIME/FREQUENCY SWITCHING



(57) Abstract

The present invention provides a new method and an apparatus for spectral envelope encoding. The invention teaches how to perform and signal compactly a time/frequency mapping of the envelope representation, and further, encode the spectral envelope data efficiently using adaptive time/frequency directional coding. The method is applicable to both natural audio coding and speech coding systems and is especially suited for coders using SBR [WO 98/57436] and other high frequency reconstruction methods.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania
AM	Armenia
AT	Austria
AU	Australia
AZ	Azerbaijan
BA	Bosnia and Herzegovina
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CU	Cuba
CZ	Czech Republic
DE	Germany
DK	Denmark
EE	Estonia
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GH	Ghana
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IL	Israel
IS	Iceland
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgyzstan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakstan
LC	Saint Lucia
LI	Liechtenstein
LK	Sri Lanka
LR	Liberia
LS	Lesotho
LT	Lithuania
LU	Luxembourg
LV	Latvia
MC	Monaco
MD	Republic of Moldova
MG	Madagascar
MK	The former Yugoslav Republic of Macedonia
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
MX	Mexico
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SG	Singapore
SI	Slovenia
SK	Slovakia
SN	Senegal
SZ	Swaziland
TD	Chad
TG	Togo
TJ	Tajikistan
TM	Turkmenistan
TR	Turkey
TT	Trinidad and Tobago
UA	Ukraine
UG	Uganda
US	United States of America
UZ	Uzbekistan
VN	Viet Nam
YU	Yugoslavia
ZW	Zimbabwe

EFFICIENT SPECTRAL ENVELOPE CODING USING VARIABLE TIME/FREQUENCY RESOLUTION AND TIME/FREQUENCY SWITCHING

TECHNICAL FIELD

The present invention relates to a new method and apparatus for efficient coding of spectral envelopes in audio coding systems. The method may be used both for natural audio coding and speech coding and is especially suited for coders using SBR [WO 98/57436] and other high frequency reconstruction methods.

BACKGROUND OF THE INVENTION

Audio source coding techniques can be divided into two classes: natural audio coding and speech coding. Natural audio coding is commonly used for music or arbitrary signals at medium bitrates, and generally offers wide audio bandwidth. Speech coders are basically limited to speech reproduction but can on the other hand be used at very low bitrates, albeit with low audio bandwidth. In both classes, the signal is generally separated into two major signal components, the "spectral envelope" and the corresponding "residual" signal. Throughout the following description, the term "spectral envelope" refers to the coarse spectral distribution of the signal in a general sense, e.g. filter coefficients in an linear prediction based coder or a set of time-frequency averages of subband samples in a subband coder. The term "residual" refers to the fine spectral distribution in a general sense, e.g. the LPC error signal or subband samples normalized using the above time-frequency averages. "Envelope data" refers to the quantized and coded spectral envelope, and "residual data" to the quantized and coded residual. At medium and high bitrates, the residual data constitutes the main part of the bitstream while the envelope data is merely a fraction. At very low bitrates, the envelope data constitutes a comparably larger part of the bitstream. Hence, it is indeed important to represent the spectral envelope compactly when using lower bitrates.

Older prior art audio coders and most speech coders use static, relatively short, time segments in the generation of envelope data to achieve good temporal resolution. However, this prevents from optimal utilisation of the frequency domain masking known from psycho-acoustics. To improve coding gain through the use of narrow filterbands with steep slopes, and still achieve good temporal resolution during transient passages, modem audio coders employ adaptive window switching, i.e. they switch time segment lengths depending on the signals statistics. Clearly a minimum usage of the short segments is a prerequisite for maximum coding gain. Unfortunately, long transition windows are needed to alter the segment lengths, limiting the switching flexibility.

The spectral envelope is a function of two variables: time and frequency. The encoding can be done by exploiting redundancy in either direction of the time/frequency plane. Generally, coding of the spectral envelope is performed in the frequency direction using delta coding (DPCM), linear prediction (LPC), or vector quantization (VQ).

SUMMARY OF THE INVENTION

The present invention provides a new method and an apparatus for spectral envelope encoding. The invention teaches how to perform and signal compactly a time/frequency mapping of the envelope representation, and further, encode the spectral envelope data efficiently using adaptive time/frequency direction coding. In the absence of transients, i.e. for quasi-stationary signals, a time/frequency grid with low temporal and high frequency resolution is used as default. In the vicinity of transients, the temporal resolution is increased at the expense of frequency resolution. The invention describes two schemes for signalling of the time and frequency resolution used. One scheme allows arbitrary selection of instantaneous resolution by explicit signalling of time segment borders and frequency resolutions, whereas the other exploits the fact that transients are separated at least by a minimum time, T_{nmin}, in order to reduce the required number of control bits: In the encoder, a transient detector decides whether the current granule contains a transient, and if so, determines the position of the onset of the transient. The position within the granule is encoded and sent to the decoder. Both the encoder and decoder share rules that specify the time/frequency distribution of the spectral envelope samples, given a certain combination of subsequent control signals, ensuring an unambiguous decoding of the envelope data. The rules can be realised as a book of tables explicitly specifying the division of the current granule in terms of samples in the time/frequency plane. The variable time/frequency resolution method is also applicable on envelope encoding based on prediction. Instead of grouping of subband samples, predictor coefficients are generated for time segments of varying lengths according to the system. Different predictor orders may be used for transient and quasi-stationary (tonal) segments.

The present invention presents a new and efficient method for scalefactor redundancy coding. A dirac pulse in the time domain transforms to a constant in the frequency domain, and a dirac in the frequency domain, i.e. a single sinusoid, corresponds to a signal with constant magnitude in the time domain. Simplified, on a short term basis, the signal shows less variations in one domain than the other. Hence, using prediction or delta coding, coding efficiency is increased if the spectral envelope is coded in either time- or frequency-direction depending on the signal characteristics.

BRIEF DESCRIPTION OF THE DRAWINGS

The present invention will now be described by way of illustrative examples, not limiting the scope or spirit of the invention, with reference to the accompanying drawings, in which:

Figs. 1a - 1b illustrate uniform respective non-uniform sampling in time of the spectral envelope.
Figs. 2a - 2c illustrate transient detector look-ahead and granule interdependency.
Figs. 3a - 3f illustrate segments with different time and frequency resolutions, and the corresponding control signals.
Fig. 4 illustrates time/frequency switched envelope coding.
Fig. 5 is a block diagram of an encoder using the envelope coding according to the invention.
Fig. 6 is a block diagram of a decoder using the envelope coding according to the invention.

DESCRIPTION OF PREFERRED EMBODIMENTS

The below-described embodiments are merely illustrative for the principles of the present invention for efficient envelope coding. It is understood that modifications and variations of the arrangements and the details described herein will be apparent to others skilled in the art. It is the intent, therefore, to be limited only by the scope of the impending patent claims and not by the specific details presented by way of description and explanation of the embodiments herein.

Generation of Envelope Data

Most audio and speech coders have in common that both envelope data and residual data are transmitted and combined during the synthesis at the decoder. Two exceptions are coders employing PNS ["Improving Audio Codecs by Noise Substitution", D. Schultz, JAES, vol. 44, no. 7/8, 1996], and coders employing SBR. In case of SBR, considering the highband, only the spectral course structure needs to be transmitted since a residual signal is reconstructed from the lowband. This puts higher demands on how to generate envelope data, in particular due to lack of "timing" information contained in the original residual signal. This problem will now be demonstrated by means of an example:

Fig. 1 shows the time/frequency representation of a musical signal where sustained chords are combined with sharp transients with mainly high frequency contents. In the lowband the chords have high energy and the transient energy is low, whereas the opposite is true in the highband. The envelope data that is generated during time intervals where transients are present is dominated by the high intermittent transient energy. At the SBR process in the decoder, the spectral envelope of the transposed signal is estimated using the same instantaneous time-/frequency resolution as used for the analysis of the original highband. An equalization of the transposed signal is then performed, based on dissimilarities in the spectral envelopes. E.g. amplification factors in an envelope adjusting filterbank are calculated as the quotients between original signal and transposed signal scalefactors. For this kind of signal, a problem arises: The transposed signal has the same chord to transient energy ratio as the lowband. The gains needed in order to adjust the transposed transients to the correct level thus cause the transposed chords to be amplified relative the original highband level for the full duration of the envelope data containing transient energy. These momentarily too loud chord fragments are perceived as pre- and post echoes to the transient, see Fig. 1a. This kind of distortion will hereinafter be referred to as "gain induced pre- and post echoes". The phenomenon can be eliminated by constantly updating the envelope data at such a high rate that the time between an update and an arbitrarily located transient is guaranteed to be short enough not to be resolved by the human hearing. However, this approach would drastically increase the amount of data to be transmitted and is thus not practical.

Therefore a new envelope data generation scheme is presented. The principal solution is to maintain a low update rate during tonal passages, which make up the majority of a typical programme material, and by means of a transient detector localize the transient positions, and update the envelope data close to the leading flanks, see Fig 1b. This eliminates gain induced pre-echoes. In order to represent the decay of the transients well, the update rate is momentarily increased in a time interval after the transient start. This eliminates gain induced post-echoes. The time segmenting during the decay is not as crucial as finding the start of the transient, as will be explained later. In order

to compensate for the smaller time steps, a lower frequency resolution can be used during the transient, keeping the data size within limits. A non-uniform sampling in time and frequency as outlined above is applicable both on subband coders and linear prediction based coders.

Some prior art coders employ variable time/frequency resolution as well. In case of subband coders, this is commonly achieved through switching of the filterbank size. Such a change in size can not take place immediately, so called transition windows are needed, and thus the update points can not be chosen freely. When using SBR, the filterbank can be designed to meet both the highest temporal and highest frequency resolution needed. Thus the varying time and frequency sampling can be obtained by grouping of the subband samples from a fixed filterbank in different ways. In other words, by keeping the filterbank size constant, high frequency resolution or high time resolution can be obtained instantaneously. In case of prediction based coders, no elaborate time/frequency resolution switching schemes are known from prior art.

Typical coders operate on a block basis, where every block represents a fixed time interval. Those blocks will be referred to as "granules". Let a granule have a length of q time quantization steps, hereinafter called "subgranules". In applications where there are non-critical delay restrictions, as in point to multipoint broadcasting, a transient detector look-ahead can be employed on the encoder side. Having this additional information, envelope data spanning across borders of granules can be comprised. This enables a more flexible selection of time/frequency resolutions, and faciliates constant bitrate operation, since parts of the payload can be moved between consecutive granules. Referring to Fig. 2, the granules are divided into eight subgranules. The transient detector operates on granules with the same timespan as the granule that overlap 50% of two consecutive granules, that is, the transient detector look-ahead is half a granule. The transient detector has detected a transient in subgranule 6 at time n-1, and a transient in subgranule 7 at time n. With these values as input to the time/frequency resolution controlling algorithms, the corresponding time/frequency grid for granule n might be as shown in Fig. 2c. As seen from the figure, subgranule 7 of the granule at time n-1 is included in the time/frequency grid of granule n. Moreover, it is possible to use an analysis by synthesis approach, i.e. having a decoder in the encoder to assess the most beneficial time/frequency sampling.

Control Signalling

In order to correctly interpret the received envelope data, the segment borders and frequency resolutions (number of coefficients or scalefactors) must be signalled. If a non-uniform sampling according to Fig. 2 is to be employed, the problem of envelope data spanning over the granule borders must be dealt with. Furthermore, the signalling must be flexible enough to cover all combinations of interest, without generating a too large amount of control data.

Theoretically, transients can occur within a granule in C combinations, ranging from no transient at all to q transients, where C is given by

$$C = \sum_{n=0}^{q} \binom{q}{n} = 2^q \qquad \text{(Eq 1)}$$

In order to signal C states, $\ln2(C) = \ln2(2^q) = q$ bits are required, corresponding to one bit per subgranule. If different frequency resolutions are to be used in the segments, even more bits might be required in order to signal the frequency resolution chosen. However, in low bitrate applications the number of control signal bits must be kept at a minimum. The first step towards an efficient signalling is to employ two time sampling modes; uniform and non-uniform sampling in time. The uniform mode is used during quasi-stationary passages, and employs high frequency resolution and relatively long time segments, both of which are predefined. Hence this mode does not require any signalling of segment borders or frequency resolutions. One bit is sufficient to signal the time sampling mode to the decoder. The non-uniform mode is used during transient passages and requires additional signalling. Two such signalling systems are proposed by the present invention.

The first system, hereinafter referred to as the "border-signalling system", uses one bit per subgranule to signal whether a segment border is present at the subgranule left border or not. Envelope data corresponding to a segment is always sent in the granule in which the segment starts. This means that the number of envelopes transmitted in a granule equals the number of left borders in the granule or the bit sum of the q border bits. The segment frequency resolutions are signalled with dynamically allocated control bits, e.g. one bit per envelope. Again, this number of bits is derived from the q border bits.

Some examples of grouping of subgranules into time segments are given in Fig. 3, where the subgranules are numbered from 000 to 111. L denotes low frequency resolution and H denotes high resolution. In the example the number of scalefactors or coefficients in a high resolution segment is assumed to be two times that of a low resolution segment. Figure 3a shows a reference system, constantly using the highest possible time and frequency resolution. The relative data matrix size is one by definition, and obviously no control signal bits are needed in this system. If no transient is present in or next to a specific granule, the granule is divided into two segments of equal length and the envelope representations are calculated using high frequency resolution. If the two envelope representations do not differ more than a certain amount, only one set of high resolution envelope data is sent. Those cases are illustrated by Figs. 3b and 3c, where the control signal "Uniform" tells that uniform sampling in time is used, and the signal "LowTime" indicates whether one or two envelopes are sent. Hence, the control signal overhead is two bits. The - symbol means that the signal is not transmitted. Figs. 3d - 3f show some cases where a transient, denoted by T, is present. The border-signalling system uses 8 bits to signal sub-granule left borders, and a varying number of bits to signal the frequency resolution within the sub-granules. Those signals are called "Borders" and "LowFreq" respectively. The "TranPos" signal is not part of this system, and will be explained later. The right border of the last segment in a granule equals the first left border in the subsequent granule. P means that the corresponding envelope data was sent in the previous granule, Fig 3f. The signalling overhead varies between 12 and 13 bits in Figs. 3d - 3f. Notice that the transient cases d and f generate the same data matrix size as the non-transient case b. Furthermore, it is possible to design a scheme that keeps the matrix size constant, if desired. For a typical programme material, the system has a performance similar to that of the reference system, at data matrix sizes of only 0.125 to 0.375 times the reference size. Hence a major data reduction is achieved when using the dynamic selection of time- and frequency resolution according to the present invention.

The second system, hereinafter referred to as the "position-signalling system", is intended for very low bitrate applications and utilizes some musical signal properties in order to reduce the number of control signal bits. As will

be shown below, many of the states described by Eq. 1 are not very likely, and would also generate too large amounts of envelope data to be practical at a limited bitrate. According to the present invention, the following simplifications can be made with little or no sacrifice of quality for practical signals:

1. Only the transient start position needs to be transmitted. The time and frequency grouping around this position can be handled by employing a set of rules in the encoder and decoder, which are based on the properties of typical transients.
2. There exists a fixed minimum time-span between consecutive transients, i.e. transients can not be arbitrarily close to one another. It is thus possible to introduce a blocking time in the transient detection/signalling system, reducing the number of states.

The minimum time-span between consecutive transients in music programme material can be estimated in the following way: In musical notation, the rhythmic "pulse" is described by a time signature expressed as a fraction A/B, where A denotes the number of "beats" per bar and $1/B$ is the type of note corresponding to one beat, for example a ¼ note, commonly referred to as a quarter note. Let t denote the tempo in Beats Per Minute (BPM). The time per note of type $1/C$ is then given by

$$T_n = (60/t)*(B/C) \text{ [s]} \qquad \text{(Eq 2)}$$

Most music pieces fall within the 70 – 160 BPM range, and in 4/4 time signature the fastest rhythmical patterns are for most practical cases made up from 1/32 or 32:nd notes. This yields a minimum time $T_{nmin} = (60/160)*(4/32) = 47$ ms. Of course lower time periods than this may occur, but such fast sequences (>21 tones per second) almost get the character of buzz and need not be fully resolved.

The necessary time resolution T_q must also be established. In some cases a transient original signal has its main energy in the highband to be reconstructed. This means that the encoded spectral envelope must carry all the "timing" information. The desired timing precision thus determines the resolution needed for encoding of leading flanks. T_q is much smaller than the minimum note period T_{nmin}, since small time deviations within the period clearly can be heard. In most cases however, the transient has significant energy in the lowband. The above described gain-induced pre-echoes must fall within the so called pre- or backward masking time T_m of the human auditory system in order to be inaudible. Hence T_q must satisfy two conditions:

$$T_q << T_{nmin} \qquad \text{(Eq 3)}$$

$$T_q < T_m \qquad \text{(Eq 4)}$$

Obviously $T_m < T_{nmin}$ (otherwise the notes would be so fast that they could not be resolved) and according to ["Modeling the Additivity of Nonsimultaneous Masking", Hearing Res., vol. 80, pp. 105-118 (1994)], T_m amounts to 10-20 ms. Since T_{nmin} is in the 50ms range, a reasonable selection of T_q according to Eq 3 results in that the second condition is also met. Of course the precision of the transient detection in the encoder and the time resolution of the analysis/synthesis filterbank must also be considered when selecting T_q.

Tracking of trailing flanks is less crucial, for several reasons: First, the note-off position has little or no effect on the perceived rhythm. Second, most instruments do not exhibit sharp trailing flanks, but rather a smooth decay curve, i.e. a well defined note-off time does not exist. Third, the post- or forward masking time is substantially longer than the pre-masking time.

According to the present invention, the above transient start information can be used for implicit signalling of segment borders and frequency resolutions immediately after/between transients. This will now be described, again referring to Fig. 3, assuming a granule length selected according to $8T_q <= T_{nmin}$, i.e. a maximum of one transient is likely to occur within a granule. In this position-signalling system the "Borders" and "LowFreq" signals are replaced by a single signal, "TranPos", consisting of three bits. When a transient is present, the position within the granule is signalled by "TranPos", see Fig. 3d – 3f. This value, in combination with the control signals of the preceding granule, determines the time/frequency grid used for the current granule. These grids are described by rules or tables that are available to both the encoder and decoder. Given the common tables and the control signals "Uniform" and either "LowTime" or "TranPos" of the current and the previous granule, unambiguous decoding of the envelope data is ensured. To put the saving obtained by the use of the position-signalling system instead of the border-signalling system into perspective, a hypothetical low bitrate envelope encoder is studied: Assume granules of length $16T_q <= T_{nmin}$, an average number of scalefactors per granule of 40 and an average number of bits per scalefactor of 3 due to lossless coding. The average number of segments in granules containing transients, n, is assumed to be 3. For transients, the signalling overheads are $B_{border} = 1 + q + n = 1 + 16 + 3 = 20$ and $B_{position} = 1 + \text{ceil}\{\ln 2(16)\} = 1 + 4 = 5$. Thus the saving is around $20 - 5 = 15$ bits, corresponding to about 5 scalefactors or 12.5 % of the envelope data, i.e. it is significant at such low bitrates.

Time/Frequency Switched Scalefactor Encoding

Utilising a time to frequency transform it can be shown that a pulse in the time domain corresponds to a flat spectrum in the frequency domain, and a "pulse" in the frequency domain, i.e. a single sinusoidal, corresponds to a quasi-stationary signal in the time domain. In other words a signal usually shows more transient properties in one domain than the other. In a spectrogram, i.e. a time/frequency matrix display, this property is evident, and can advantageously be used when coding spectral envelopes.

A tonal stationary signal can have a very sparse spectrum not suitable for delta coding in the frequency-direction, but well suited for delta coding in the time-direction, and vice versa. This is displayed in Fig. 4. Throughout the following description a vector of scale factors calculated at time n_0 represents the spectral envelope

$$Y(k,n_0)=[a_1, a_2, a_3, \ldots, a_k, \ldots, a_N], \qquad \text{(Eq 5)}$$

where $a_1 \ldots a_N$ are the amplitude values for different frequencies. Common practice is to code the difference between adjacent values in the frequency-direction at a given time, which yields:

$$D(k,n_0)=[a_2-a_1, a_3-a_2, \ldots, a_N-a_{(N-1)}]. \qquad \text{(Eq 6)}$$

In order to be able to decode this, the start value a_1 needs to be transmitted. As stated above this delta-coding scheme can prove to be most inefficient if the spectrum only contains a few stationary tones. This can result in a delta coding yielding a higher bit rate than regular PCM coding. In order to deal with this problem, a time/frequency switching method, hereinafter referred to as T/F-coding, is proposed: The scalefactors are quantized and coded both in the time- and frequency-direction. For both cases, the required number of bits is calculated for a given coding error, or the error is calculated for a given number of bits. Based upon this, the most beneficial coding direction is selected.

As an example, DPCM and Huffman redundancy coding can be used. Two vectors are calculated, D_f and D_t:

$$D_f(k,n_0)=[a_2-a_1,a_3-a_2,\ldots,a_N-a_{(N-1)}], \qquad \text{(Eq 7)}$$

$$D_t(k,n_0)=[a_1(n_0)-a_1(n_0-1),a_2(n_0)-a_2(n_0-1),\ldots,a_N(n_0)-a_N(n_0-1)] \qquad \text{(Eq 8)}$$

The corresponding Huffman tables, one for the frequency direction and one for the time direction, state the number of bits required in order to code the vectors. The coded vector requiring the least number of bits to code represents the preferable coding direction. The tables may initially be generated using some minimum distance as a time/frequency switching criterion.

Start values are transmitted whenever the spectral envelope is coded in the frequency direction but not when coded in the time direction since they are available at the decoder, through the previous envelope. The proposed algorithm also require extra information to be transmitted, namely a time/frequency flag indicating in which direction the spectral envelope was coded. The T/F algorithm can advantageously be used with several different coding schemes of the scalefactor-envelope representation apart from DPCM and Huffman, such as ADPCM, LPC and vector quantisation. The proposed T/F algorithm gives significant bitrate-reduction for the spectral-envelope data, up to around 20% reduction compared to commonly used delta-coding techniques. If the number of scalefactors per octave is constant, it is possible to delta code on an octave basis instead of delta coding of adjacent scale factors.

Practical implementations

An example of the encoder side of the invention is shown in Fig. 5. The analogue input signal is fed to an A/D-converter 501, forming a digital signal. The digital audio signal is fed to a perceptual audio encoder 502, where source coding is performed. In addition, the digital signal is fed to a transient detector 503 and to an analysis filterbank 504, which splits the signal into its spectral equivalents (subband signals). The transient detector could operate on the subband signals from the analysis bank, but for generality purposes it is here assumed to operate on the digital time domain samples directly. The transient detector divides the signal into granules and determines, according to the invention, whether subgranules within the granules is to be flagged as transient. This information is sent to the envelope grouping block 505, which specifies the time/frequency grid to be used for the current granule. According to the grid, the block combines the uniform sampled subband signals, to form the non-uniform sampled envelope values. As an example, these values might be the average or maximum energy for the subband samples combined. The envelope values are, together with the grouping information, fed to the envelope encoder block 506. This block decides in which direction (time or frequency) to encode the envelope values. The resulting signals, the output from the audio encoder, the wideband envelope information, and the control signals are fed to the multiplexer 507, forming a serial bitstream that is transmitted or stored.

The decoder side of the invention is shown in Fig. 6. The demultiplexer 601 restores the signals and feeds the appropriate part to an audio decoder 602, which produces a low band digital audio signal. The envelope information is fed from the demultiplexer to the envelope decoding block 603, which, by use of control data, determines in which direction the current envelope are coded and decodes the data. The low band signal from the audio decoder is routed to the transposition module 604, which generates a replicated high band signal consisting of one or several

harmonics from the low band signal. The high band signal is fed to an analysis filterbank 606, which is of the same type as on the encoder side. The subband signals are combined in the scalefactor grouping unit 607. By use of control data from the demultiplexer, the same type of combination and time/frequency distribution of the subband samples is adopted as on the encoder side. The envelope information from the demultiplexer and the information from the scalefactor grouping unit is processed in the gain control module 608. The module computes gain factors to be applied to the subband samples before recombination in the synthesis filterbank block 609. The output from the synthesis filterbank is thus an envelope adjusted high band audio signal. This signal is added to the output from the delay unit 605, which is fed with the low band audio signal. The delay compensates for the processing time of the high band signal. Finally, the obtained digital wideband signal is converted to an analogue audio signal in the digital to analogue converter 610.

CLAIMS

1. A method for spectral envelope coding in a source coding system where said system comprises an encoder representing all operations performed prior to storage or transmission, and a decoder representing all operations performed after storage or transmission, **characterised by**:

at said encoder, perform a statistical analysis of the input signal,

based on the outcome of said analysis, select the instantaneous time and frequency resolution to be used in the spectral envelope representation,

using said resolution, generate data representing said spectral envelope,

transmit said data together with a control signal describing said resolution, and

at said decoder, using said control signal and said data in the synthesis of the output signal.

2. A method according to claim 1, **characterised in** that said instantaneous time and frequency resolution is obtained by grouping of elements in a time/frequency representation of said input signal, and calculating a scalefactor for every one of said groups.

3. A method according to claim 2, **characterised in** that said time/frequency representation is generated by a filterbank.

4. A method according to claim 3, **characterised in** that said filterbank is of fixed size.

5. A method according to claim 1, **characterised in** that said data is generated by a linear predictor.

6. A method according to claim 1, **characterised in** that said analysis employs a transient detector.

7. A method according to claim 6, **characterised in** that said instantaneous resolution is switched from a default combination of higher frequency resolution and lower time resolution to a combination of lower frequency resolution and higher time resolution at the onset of a transient.

8. A method according to claim 1, **characterised in** that said control signal describes positions within a granule of constant update rate, generated by said analysis, and said instantaneous resolution is chosen based on the positions within current and neighbouring granules, by the use of rules available to both said encoder and said decoder.

9. A method according to claim 8, **characterised in** that at most one position per granule is signalled.

10. A method according to claim 1, **characterised in** that said control signal describes borders within a granule of constant update rate, said instantaneous resolution is signalled once per border, and that one set of data is sent per border within a granule.

11. A method according to claim 2, **characterised in** that said scalefactors are coded both in the time and frequency direction, the momentarily most beneficial direction is determined, said most beneficial direction is used for said transmission.

12. A method according to claim 11, **characterised in** that the direction which generates the least coding error for a given number of bits is chosen.

13. A method according to claim 11, **characterised in** that the direction which generates the least number of bits for a given coding error is chosen.

14. A method according to claim 13, **characterised in** that lossless coding is employed and separate tables are used for said time and frequency directions, in particular where said tables are used for selection of coding direction.

15. An apparatus for encoding of a spectral envelope of a signal to be decoded by a decoder, **characterised by**:
means for performing a statistical analysis of the input signal,
means for selection of the instantaneous time and frequency resolution to be used in a spectral envelope representation of said input signal, based on the outcome of said analysis,
means for generation of data representing said spectral envelope, using said resolution, and
means for transmission of said data together with a control signal describing said resolution.

16. An apparatus for decoding of a spectral envelope of a signal encoded by an encoder, **characterised by**:
means for interpretation of a received control signal in order to determine the instantaneous time and frequency resolution used in a spectral envelope representation of an encoded signal,
means for decoding of received envelope data based on said spectral envelope representation, using said control signal, and
means for using said decoded envelope data in the synthesis of the output signal.

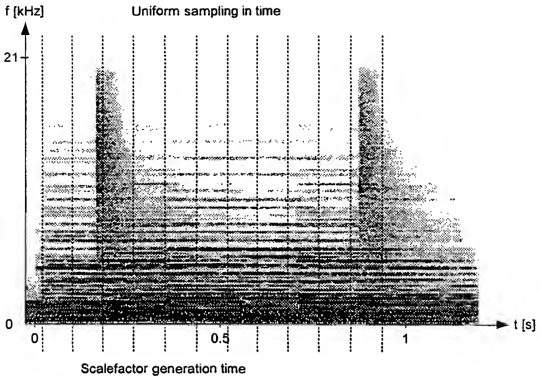

f [kHz]
Uniform sampling in time
21
0
0
0.5
1
t [s]
Scalefactor generation time

Fig. 1a

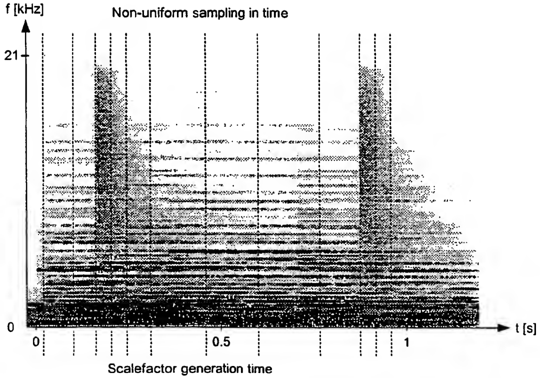

f [kHz]
Non-uniform sampling in time
21
0
0
0.5
1
t [s]
Scalefactor generation time

Fig. 1b

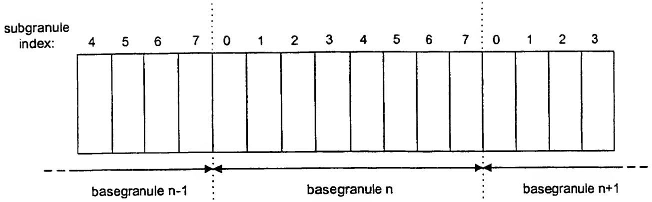


Fig. 2a

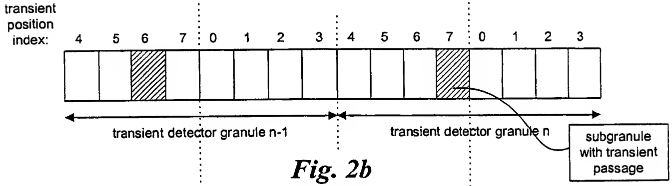


Fig. 2b

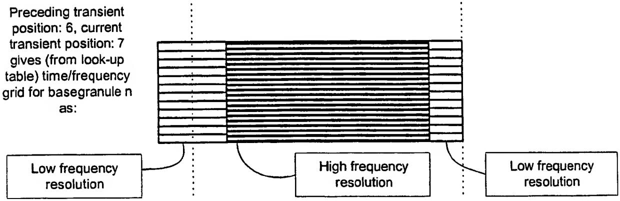


Fig. 2c

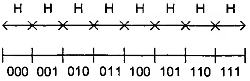


Fig. 3a

	Reference-System
Matrix Rel Size	1
# Control Bits	0

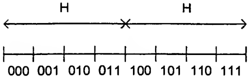


Fig. 3b

	Border-System	Position-System
Matrix Rel Size	0.25	
Uniform	1	
LowTime	0	
Borders	-	N/A
LowFreq	-	N/A
TranPos	N/A	-
# Control Bits	2	2

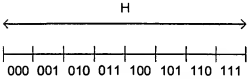


Fig. 3c

	Border-System	Position-System
Matrix Rel Size	0.125	
Uniform	1	
LowTime	1	
Borders	-	N/A
LowFreq	-	N/A
TranPos	N/A	-
# Control Bits	2	2

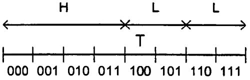


Fig. 3d

	Border-System	Position-System
Matrix Rel Size	0.25	
Uniform	0	
LowTime	-	
Borders	10001010	N/A
LowFreq	011	N/A
TranPos	N/A	100
# Control Bits	12	4

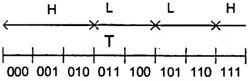


Fig. 3e

	Border-System	Position-System
Matrix Rel Size	0.375	
Uniform	0	
LowTime	-	
Borders	10010101	N/A
LowFreq	0110	N/A
TranPos	N/A	011
# Control Bits	13	4

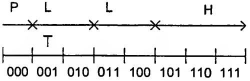


Fig. 3f

	Border-System	Position-System
Matrix Rel Size	0.25	
Uniform	0	
LowTime	-	
Borders	01010100	N/A
LowFreq	110	N/A
TranPos	N/A	001
# Control Bits	12	4

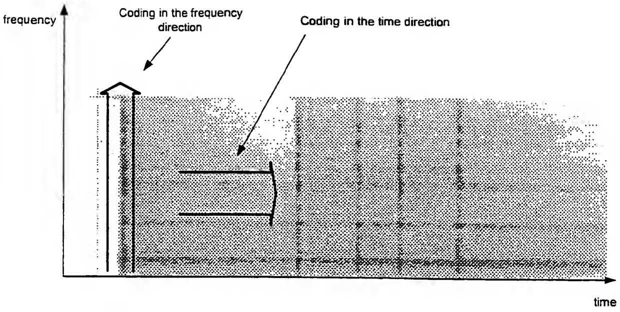

frequency
Coding in the frequency direction
Coding in the time direction
time

Fig. 4

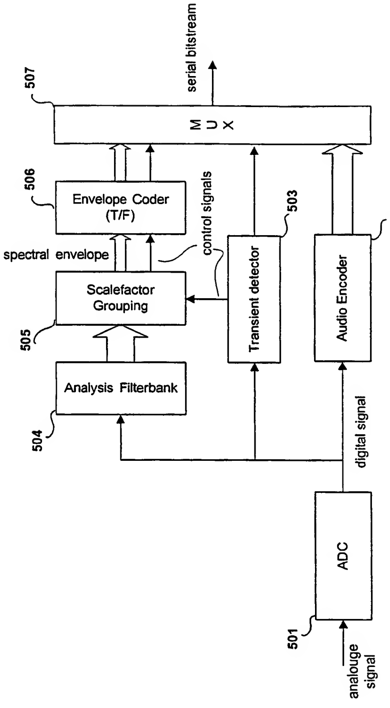

serial bitstream
507
MUX
506
Envelope Coder
(T/F)
control signals
503
502
spectral envelope
Scalefactor
Grouping
Transient detector
Audio Encoder
505
Analysis Filterbank
504
digital signal
ADC
501
analouge
signal

Fig. 5

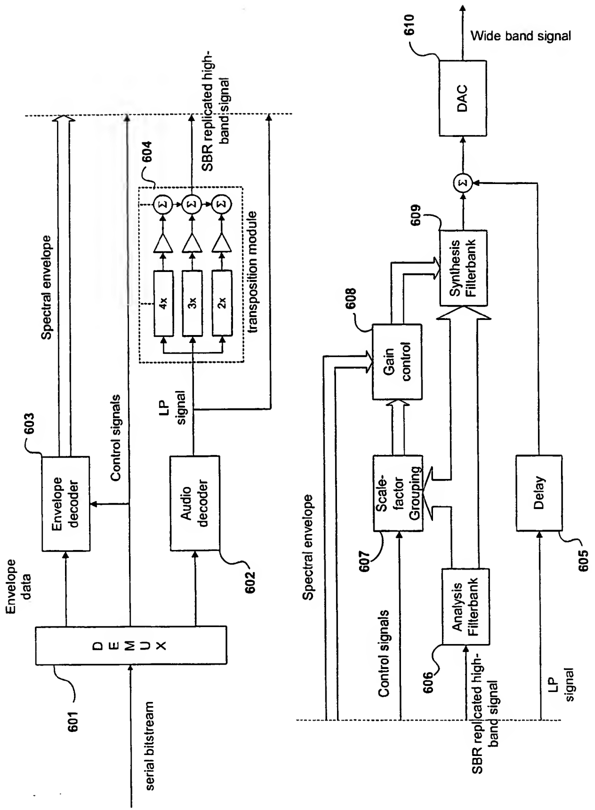

serial bitstream
601
DEMUX
Envelope data
Envelope decoder
603
Control signals
Spectral envelope
Audio decoder
602
LP signal
4x
3x
2x
transposition module
604
SBR replicated high-band signal
Spectral envelope
Control signals
SBR replicated high-band signal
LP signal
Analysis Filterbank
606
Scale-factor Grouping
607
Gain control
608
Delay
605
Synthesis Filterbank
609
DAC
610
Wide band signal

Fig. 6

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : G10L 19/00 | A3

(11) International Publication Number: WO 00/45378

(43) International Publication Date: 3 August 2000 (03.08.00)

(21) International Application Number: PCT/SE00/00158

(22) International Filing Date: 26 January 2000 (26.01.00)

(30) Priority Data:
9900256-0 27 January 1999 (27.01.99) SE
9903552-9 1 October 1999 (01.10.99) SE

(71)(72) Applicant and Inventor: LILJERYD, Lars, Gustaf [SE/SE]; Vintervägen 19, S-171 34 Solna (SE).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: KJÖRLING, Kristofer [SE/SE]; Lostigen 10, S-170 75 Solna (SE). EKSTRAND, Per [SE/SE]; Renstiernas Gata 29, S-116 31 Stockholm (SE). HENN, Fredrik [SE/SE]; Ritarvägen 14, S-168 31 Bromma (SE).

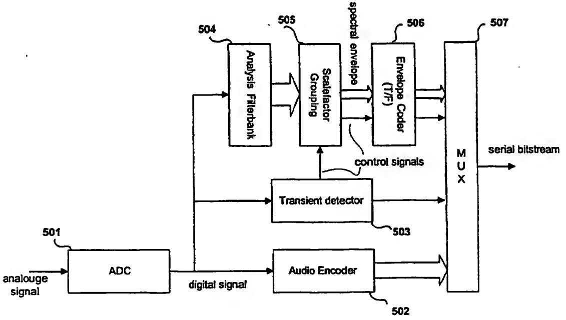
(74) Agents: ÖRTENBLAD, Bertil et al.; Noréns Patentbyrå AB, Box 10198, S-100 55 Stockholm (SE).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(88) Date of publication of the international search report: 16 November 2000 (16.11.00)

(54) Title: EFFICIENT SPECTRAL ENVELOPE CODING USING VARIABLE TIME/FREQUENCY RESOLUTION AND TIME/FREQUENCY SWITCHING



(57) Abstract

The present invention provides a new method and an apparatus for spectral envelope encoding. The invention teaches how to perform and signal compactly a time/frequency mapping of the envelope representation, and further, encode the spectral envelope data efficiently using adaptive time/frequency directional coding. The method is applicable to both natural audio coding and speech coding systems and is especially suited for coders using SBR [WO 98/57436] and other high frequency reconstruction methods.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania
AM	Armenia
AT	Austria
AU	Australia
AZ	Azerbaijan
BA	Bosnia and Herzegovina
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CU	Cuba
CZ	Czech Republic
DE	Germany
DK	Denmark
EE	Estonia
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GH	Ghana
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IL	Israel
IS	Iceland
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgyzstan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakstan
LC	Saint Lucia
LI	Liechtenstein
LK	Sri Lanka
LR	Liberia
LS	Lesotho
LT	Lithuania
LU	Luxembourg
LV	Latvia
MC	Monaco
MD	Republic of Moldova
MG	Madagascar
MK	The former Yugoslav Republic of Macedonia
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
MX	Mexico
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SG	Singapore
SI	Slovenia
SK	Slovakia
SN	Senegal
SZ	Swaziland
TD	Chad
TG	Togo
TJ	Tajikistan
TM	Turkmenistan
TR	Turkey
TT	Trinidad and Tobago
UA	Ukraine
UG	Uganda
US	United States of America
UZ	Uzbekistan
VN	Viet Nam
YU	Yugoslavia
ZW	Zimbabwe

INTERNATIONAL SEARCH REPORT

International application No.
PCT/SE 00/00158

A. CLASSIFICATION OF SUBJECT MATTER

IPC7: G10L 19/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC7: G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

SE,DK,FI,NO classes as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	PRINCEN, J. ET AL "Audio coding with signal adaptive filterbanks" In: 1995 International Conference on Acoustics, Speech and Signal Processing, ICASSP-95, vol. 5, pages 3071 - 3074, see the whole document --	1-16
X	US 5852806 A (JAMES DAVID JOHNSTON ET AL), 22 December 1998 (22.12.98), column 2, line 3 - column 3, line 4; column 4, line 15 - column 6, line 7 --	1-16

[X] Further documents are listed in the continuation of Box C. [X] See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" erlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
12 July 2000	19-07-2000
Name and mailing address of the ISA/ Swedish Patent Office Box 5055, S-102 42 STOCKHOLM	Authorized officer PEDER GJERVALDSAETER/EE

INTERNATIONAL SEARCH REPORT

International application No.
PCT/SE 00/00158

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	BOSI, M. ET AL "Time versus Frequency Resolution in a Low-Rate, High Quality Audio Transform Coder" In: 1991 IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics, Final Program and Paper Summaries, pages 0_81 - 0_82, see the whole document --	1-16
A	US 5581653 A (CRAIG C. TODD), 3 December 1996 (03.12.96), column 4, line 19 - column 5, line 6, abstract --	1-16
A	US 5737718 A (KYOYA TSUTSUI), 7 April 1998 (07.04.98), column 3, line 8 - column 4, line 63 --	1-16
A	WO 98/57436 A2 (LILJERYD, LARS GUSTAF), 17 December 1998 (17.12.98), column 2, line 18 - column 5, line 10, abstract --	1-16
A	US 5504832 A (TETSU TAGUCHI), 2 April 1996 (02.04.96), column 1, line 65 - column 4, line 7 -- --------	1-16

INTERNATIONAL SEARCH REPORT
Information on patent family members

02/12/99

International application No.
PCT/SE 00/00158

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 5852806 A	22/12/98	CA 2199070 A	19/09/97
		EP 0797313 A	24/09/97
		JP 10039897 A	13/02/98
US 5581653 A	03/12/96	AT 147910 T	15/02/97
		AU 685505 B	22/01/98
		AU 7676594 A	22/03/95
		CA 2167527 A	09/03/95
		DE 69401517 D,T	12/06/97
		DK 716787 T	23/06/97
		EP 0716787 A,B	19/06/96
		SE 0716787 T3	
		ES 2097061 T	16/03/97
		JP 9502314 T	04/03/97
		SG 48278 A	17/04/98
		WO 9506984 A	09/03/95
US 5737718 A	07/04/98	JP 7336232 A	22/12/95
WO 98/57436 A2	17/12/98	NONE	
US 5504832 A	02/04/96	AU 657184 B	02/03/95
		AU 3019692 A	01/07/93
		CA 2085384 A,C	25/06/93
		CA 2193345 A	25/06/93
		JP 5173599 A	13/07/93